

Watson & Crickの論文：“The Structure of DNA”の語彙分析 —JACET8000を基準にした分類と学生アンケートとの対比

Analyzing the vocabulary of “The Structure of DNA” by Watson & Crick
—on the basis of JACET8000, contrasted with the students' questionnaire

普遍英語 石井 有美

1. はじめに

千葉大学の普遍英語で科学関連のテキストを使っているうちに、DNA研究のWatson & Crickの論文に関心を持ち、1953年の論文を千葉大学図書館で入手してもらった。1953年といえば50年以上も前のことである。一般にDNAが知れ渡りようになったのは、遺伝子操作の食品や食の安全性についての議論が盛んになった頃だろうか。DNAの研究自体はかなり以前からなされていたわけである。サイエンスの論文は専門語彙に限られていて、それを把握すれば高等学校までの学習語彙で読めるとかねて聞いていたが、果たしてどのような語彙が用いられているだろうか。

1.1 なぜ50年も前の論文を扱うのか

なぜ50年も前の論文を扱うのかという指摘があった。遺伝子操作の大豆食品が話題になったのは、この10～20年であったと思うが、DNAの論文の年代を正確に知っているのは生命科学の専門家ぐらいであろう。筆者も1953年という年代については、30年ぶりの同窓会でDNAの博士号をもつ友人に再会して初めて知ったのである。日頃使っている教科書にDNAが何度か現れていたのも、その意味ではDNAが身近なテーマであり、ひとりでも興味をもって語彙分析の対象として選んでみた。2008年のテーマとしては、春にはiPS細胞の話題があり、秋には日本人が4人もノーベル賞を受賞したという記録的な年でもあるので、その分野の論文でも入手するのがよいと思うのだが、この分野は次回の課題にしたいと思う。

1.2 サイエンス教科書に現れるDNA

普遍英語でサイエンスのテーマのテキストを使っている。例えばScience World vol 2の第1課がDNAである。1997年クローン羊のドリーが生まれて、動物のクローニングについて議論が盛んだった。ノーベル賞受賞者でカリフォルニアSalk InstituteのRenato Dulbecco博士が、人間のDNAの完全なマッピングを提案して完成したのが2000年6月とあったことがテキストで紹介されている。さらにテキストでは遺伝子による栽培の現状について知ることができる。さらにScience Spectrumという他のテキストでは、遺伝子ドーピングのテーマがあった。遺伝子ドーピングは探知できないのだそうだ。DNAのテーマは50年前であると同時に、扱い方による可能性の広さという意味では無限に未来的でもあ

り、古くて新しいテーマ、時代を超えているのではと考える。本質的なテーマは時代を超えているものだ。50年前ということを特に意識することなく、論文の原文を入手して読んでみたいと思ったのがきっかけであった。普遍英語で副題が学術の場合に英語を指導する際に、論文の語彙など提示できれば読解指導だけでなく、リスニングにもスピーキングにも、さらにライティングにはもちろんのこと大いに役立つことと思う。

1.3 学生のアンケート

本年度、Science Spectrumという教科書を使っている学生にアンケートを行なってみた。1年生ブロック別園芸学部32名のクラスである。

まずワトソンとクリックを知っているかどうかについては、以前から知っていた(21名、66%) 名前は聞いたことがある(2名6%) 全く知らなかった(9名28%)

2重らせん構造について、以前から知っていた(25名78%) 聞いたことがある(2名6%) 知らなかった(5名16%) さらにDNAについての興味では、とても興味がある(6名19%) ある程度興味がある(19名59%) 大して興味がない(7名22%)

ワトソンとクリックや二重らせんについては、7~8割の学生が知っているわけだが、大学受験の語彙で科学論文を読みこなせると思うかどうかについては、思うと答えた学生は3名9%、29名91%は思わないと答えている。

受験の英語で論文語彙を何パーセントぐらいカバーしていると思いますかについては20~30%と答えたのは9名28%、30~40%が7名22%、40~50%は同じく7名22%、50~60%は1名3%、60~70%は5名16%、70~80%はゼロ、80~90%は3名9%であった。72%の学生が50%以下と回答している。

このアンケート結果をふまえて、実際の分析結果と対比させてみたい。学生の意識とのずれが明確にでるだろうか。

1.4 分析の方法

昨今はコーパスを使った大規模な語彙分析が盛んな時代で、生命科学の語彙については、例えば2007年のJACET大会では「生命科学分野で使用される英語語彙の分析：頻出語彙リストを使用した複数コーパスの比較」という研究発表が萩原明子氏によってなされた。本稿は、こうした大規模な分析ではなく、本文8ページほどの論文を手作業で行なったものである。語彙分析としては、全体語数の把握、内容語と機能語にわけ、内容語についてどのような単語が最も頻繁に用いられているか、語の頻度数をだす、内容語から専門用語を抽出する、専門用語と専門用語以外の語についてJACET8000に従って分類してみる、以上の手順、プロセスで論文の語彙を検討してみたい。

2 全体語数の把握

Cold Spring Harbor symposium on quantitative biology 1953; 18:123-31のThe Structure of DNAの論文は、本文8ページと参考文献1ページの文章であるが、全体の語数をまず把握しておきたい。ただし、参考文献と本文中の写真や表の説明部分の単語は含んでいない。さらに、算用数字と年代を表す数字も全体語数には加えなかった。ただし例えば

算用数字の3は除外しても、スペルでthreeと書かれている場合には、全体語数に加えて数えてあるので語彙分類にもっている。それらを除いて、内容語と機能語をあわせた総数をだしてみた。全体の語数は次の通りである。

P123では404語、p124 501語、p125 355語、p126 614語、p127 959語、p128 979語、p129 987語、p130 331語。

これらを合計すると、5130語となる。5130語を基準にして検討したい。

3 内容語と機能語に分類する

3.1 内容語と機能語の定義

内容語は、物、物事、資質、状態や行動を表す語で、その意味は文法を表すものではなく、意味を表す語である。名詞、形容詞、本動詞、副詞、数詞、これらを内容語という。内容語は文全体の意味に欠かせないもので大きな影響を与える。一語でも意味をなすことができ、自立した語といえる。

これに対して機能語は文法的な機能を表す語である。その語1語では意味をなすことができないで、他の語に依存して存在する。接続詞、限定詞（名詞の前に位置してその名詞を限定する機能語）、人称代名詞、疑問詞、関係代名詞がふくまれる。さらにTaylor (1979)によれば、be, seem, appear, becomeなどは動詞であるが内容語でなく機能語と分類される。

本稿のWatson & Crickにおいては、seem は7回、appearは4回用いられていた。さらに、助動詞（can, may, mustなど）、still, yetなどの時制を示すtense marker, it やthere, 比較級、最上級を表すmore やmost, 否定のnotも機能語に分類する。冠詞、前置詞、接続詞、間投詞、これらも機能語とみなす。人称代名詞や疑問詞、否定辞も文全体の意味に欠かせないものではという指摘がある。否定語のnotは全体で9回、noは5回しか用いられていなかった。8ページに渡る論文の中で9回や5回とは、否定辞の少なさが驚きではあった。確かに肯定文か否定文かは大きな意味の違いなので、除外することはできないと思うが、ここでは頻度が少なかったこともありTaylorの分類に従って行ってみた。

例えば、They analyzed the vocabulary of the thesis. においては、内容語は analyze, vocabulary, thesis で、機能語はthey, the, ofである。リスニングでこの文を聴いた場合に、内容語は強く発音されるが、機能語は低く発音されて、意味に大きな影響を与えない語と見なされるわけである。

3.2 内容語と機能語の総数と異なり語の定義

全体語数には算用数字や年代を表す数字を除外してあるが、総数5130のうち、内容語は2442語、機能語は2688語であった。内容語について、どのような語彙が用いられているのか、その異なり語と頻度数について以下のセクションで述べる。機能語は本稿では分析していない。

異なり語とはdifferent wordsである。例えばThey analyzed the vocabulary of the thesis において、総語数は7語、内容語はanalyze, vocabulary, thesesの3語、その他のthey, the, of, theの4語は機能語である。さらにtheが2回用いられているので、異なり語数と

しては6語となる。同じ単語で品詞が違う場合、例えばHe dreamed a happy dream. では、異なり語は、he, dream, a, happyの4語である。異なり語では、品詞が違っても同じ語とみなすことができる。

4 語彙の頻度数の分類とその基準

Watson & Crickの論文で、1ページ分の参考文献は分析対象にしなかった。本文部分について、こういった種類の内容語が何回用いられているか、調べてみた。以下に頻度数順に各単語を列挙してみる。数字は頻度数を示す。頻度数が高ければ重要な語といえるのかどうかであるが、必ずしも相関性はないかもしれないが、頻度数が高いということは、少なくとも学習するに値する語、教えるに値する語であるということ是可以である。

当初は、品詞を考慮するつもりだったが、同じスペルの単語の場合は品詞が違っても同一語として数えることにした。またJACET8000の分類に従って、動詞の活用形の変化形については同じ語として数えることにした。ing形についても原形にもどして分類した。現在形は用いられていないが、過去形だけ用いられている動詞があって、例えば頻度数1のbroughtやbeganでは過去形ではなく原形で表記した。形容詞では、high 2回、higher 4回の場合は、あわせてhigh 6回と分類した。数字の算用数字は全体語数から除外してあるが、thirtyのようにアルファベットになっている場合には、語数として数えてあるのでリストに載っている。

4.1 頻度数20以上の単語

最も頻度数が高いのは、頻度数55回のstructureという単語だった。次が、52回でDNAである。第3位は48回のchainである。第4位は44回のbase、第5位の頻度数34はAである。ここで、AはDNAのタイプを表す。A型とB型がある。33回はpairだが、pairについては名詞として15回、動詞9回、pairingとして9回である。32回はformで動詞として16回、名詞として16回である。29回がone、24回はcytosine、model、23回fibre、と続く。

注：cytosine キトシン（核酸の成分）

4.2 頻度数10から19の語は25語

19 bond suggest 18 coil 16 hydrogen 15 only
 13 occur have (hasを含む) sequence
 12 figure group show (show6回、shown6回) single X-ray
 11 difficulty phosphate replication specific sugar together
 10 backbone helical helice nucleotide polynucleotide way

註：helical 螺旋（らせん）の意味、spiralでなくhelicalを用いている

nucleotideはヌクレオチド、核酸の構成成分

phosphate 磷酸塩 磷酸エステル

polynucleotide ではpoly-は、～多、～重、～複で重合体（ポリマー）の意味

replication 生物化学では遺伝子の複製

4.3 頻度数4から9の単語

頻度数9は9語である。

adenine, complementary, crystallographic, first, guanine, hold (hold 3回、held 3回、holding 3回) mythl, thus, turn

注：adenine アデニンは核酸に含まれる塩基のひとつ

guanine は核酸構成物質のひとつ

頻度数8の語は10語である。

approximately believe evidence feature find join pattern protein
specificity thymine

頻度数7は13語である。

B data end genetic helix limit link living PH
same strand unit Wilkins

注：BはDNAタイプを表す。A型とB型がある。

Wilkins はイギリスの生物物理学者。1962年にはWatson & Crickとともにノーベル医学・生理学賞を受賞

頻度数6の語は20語

acid apart atom consist distance high instead involve itself part per-
cent phage point possible regular result
scheme some task tautomeric

頻度数5は23語

adenine assume break chemical direction fact fit Franklin give
know large length number obtain obvious order position
postulate process purine reflexion separate side

頻度数4は50語

account allow always arrange both carry case cell change come different
either emphasize high highly irregular likely long make many mechanism
molecular nature neighbor new occur organism paranemic perpendicular
plectonemic present propose pyrimidine regularly right secondly self simple
space strong structural thirty time total unlikely untwist very virus
weight Wyatt

4.4 頻度数3の単語は72語 スラッシュは10語ごとの区切れを表す

alone also characteristic chromosome completely complimentary common
conclusion considerably contain / correspond Crick curve diameter difference
discuss evidence example experimental follow / free general genetical given
glucosidic hand hydroxyl imagine indicate least / length like lower
manner material meridian moment much necessary necessity / opposite oc-
casionally outside particular possibility precise preliminary present proba-

ble problem / produce reason related round second serve similar start stereochemical strongly / study support take tangling technique template therefore think three unwind / upper Watson

注 likeは、前置詞1回、動詞1回、名詞1回で合計3回である。

thinkはthoughtとして3回用いられている。

4.5 頻度数2の語は182語

スラッシュは10語の区切れ、ダブルスラッシュは50語を表す

accessible accomplish adenine again all allow alternate analysis analytical argument / arrangement assumption Astbury attach available avoid back bacteriophage bond build / common component concentration condition confirm consequence contact correct co-worker currently / decrease demand describe detail detailed difficult dimension discussion distinct done / drawn due duplicate duplication electron end enzyme equal essential establish // every exact exert exist explain fact fall far favor formation / four Fraser function further go Gosling grow Gulland head histone / humidity hypothesis implication imply important impose impossible inference information initial / initially inside intact interaction interpretation involve isolation Jordan lattice left / length less line link linkage long main mainly mean mechanism // measurement monomer moreover mould mutation none note nothing observe offer / open originally pack particle perhaps phage plausible polymerization precursor presence / proof propose protamine prove push rather ratio regard regular regularity / regularly relative relationship release remain replicate reproduction require restrictive restriction / reversible ribbon-like rigid rotation roughly run see self-duplication sense small // solution soon source space spite spite strengthen successive suspect / synthesis T tentative thin third thousand titration titratable twist type / unable unattractive uncoil use usually various vary view viscosity water / well work

注 Tはthymine チミンのこと

4.6 頻度数1の単語は456語。

スラッシュは10語の区切れ、ダブルスラッシュは50語を示す。

able absent acceptable accessible accompany accord accumulate achieve act active / actual addition adjacent alkali almost already altar alternative angle angstrom / anomalous answer appearance apposition approach arise around arrangement artifact ascertain / ask aspect assess assist asymmetrical attach attempt atomic author automatically / average axially bacterium bundle bear begin bend better biographically biologically // birefringence birefringent both bridge bring calf cause capable capacity carrier / certain center chance Chargaff circumference clear clearly clue Cohen coincide /

college combine come commonly compact companion compensate complement
 component compound / conception conclusion configuration connection contain con-
 tract control course crystal crystalline / crystallize cytologist describe diffusion
 dichroism diagram draw danger deduce define // definite Delbruck density de-
 polymerization description diagram dichroism di-ester differ diffraction / diffusion
 dimensional direct direction directly disappearance displaced dispose disproof dis-
 tant / distinguish double down draw dry dyad dylomerization early easily easy
 / effectively eighth element eleven emerge employ energy enough enthusiastic
 equally / equatorial equivalent especially establish estimate estimation even exist
 existence express // extract extraction extremely face facilitate fact fairly far
 fashion favored / fibulous fibrous fifth fifty final find finish fixed flat flow /
 fold force formidable formula formulate four fulfill fundamental get generally /
 geometrical glucoside good Gosling grasp Harshey held hope Huskin idea /
 ignore illustrate immediately imperative importance important improve inaccessi-
 ble including increase // independent indirect individual inference influence in-
 frared instead insuperable interest interesting / interlace intermediary internucle-
 otide interval interpreted intertwine intriguing introduce irrespective isolated /
 isolation ionizable ionization justify kind King's latter lead level lie / light lo-
 cated location London long low macroscopic make-up maintain manipulation /
 Manton many material matter mean measured member merely meridinal merit
 // method microscope microscopy million mind Muller MW natural Nature near-
 ly / necessarily never nevertheless newly next nitrogenous normal normally now
 nucleic / nucleoprotamine nucleotic number objection observation occurrence oppose
 ordered orientation original / otherwise outline over owing para-crystalline parallel
 paramenic parental particle particular / partner pass Pauline penny period perpet-
 uate phage phosphorous phosphoryl photograph // physical physical-chemical
 physiology picture pile pitch place plectonemically point polypeptide / precisely
 preparation probably proceed profitably proof proportion proposition proximity
 pseudo-hexagonal / purify puzzling pull purify pyrimidine quite qualitative Ran-
 doll raise range / rapid reasonable recently recourse reduce reexamine regularly
 rejoin relation relatively / relevant repeat replace requirement residue resist re-
 solve resolution resume role // room roughly salt satisfactorily say section second-
 ary sedimentation see self-reproduction / separate series seventy-five shape shift
 shorten significant simplicity simply Siegal / Singer six sixth slip small sodi-
 um solution sometimes sort spacing / specifically special sperm spontaneously
 stable stack stage state steric stereochemically / straight strain striking strong
 strictly structural subject suggestion sum super // superficially superfluous sur-

mount surprising surprisingly symmetrically symposium synthesize table tabulate / take tangled tautometric Taylor tend Tennent term themselves therefore thread / through theory thymus tilted topological transition translator treat try ultraviolet / unable unambiguously unbranch undegrade unduplicated uniform uninterpreted union unity up / usual untreaded van der Waal variance variability variety viewpoint Vibrandt volume well-defined // whilst width Williams worthwhile write wrong yield zipper-like

注 MWはmolecular weight である。

頻度数1の単語は以上のように458語である。

5 内容語の種類の数と延べ語数

以上の4.1から4.6のまとめとして、内容語の種類、異なり語数と延べ語数をまとめてみると、以下の表のようになる。

頻度数	55回	52	48	44	34	33	32	29
異なり語	1	1	1	1	1	1	1	1
延べ語数	55	52	48	44	34	33	32	29
頻度数	24	23	19	18	16	15		
異なり語	2	1	2	1	1	1		
延べ語数	48	23	38	18	16	15		
頻度数	13	12	11	10	9	8	7	
異なり語	3	5	6	6	9	10	13	
延べ語数	39	60	66	60	81	80	91	
頻度数	6	5	4	3	2	1		
異なり語	20	23	50	72	182	458		
延べ語数	120	115	200	216	364	458		

これらを合計すると、内容語の種類として、異なり語は873語で、延べ語数は2435語となった。

6 専門用語の語彙一覧

内容語は明らかになったが、その中に生命科学用語がどの程度含まれているかを調べた。筆者の判断によって判別したものだが、頻度数順にどのような専門用語が用いられているのか、その異なり語の種類を明らかにしたい。頻度数の高かった語のうち、chainはJAC-

ET8000リストで1792に分類されている。そのほかbase 519, bond 2502, sequence 3112, coil 5825, structure 1235, model 928, virus2111, element 1237, particle 3500, hydrogen4469に分類されている。これらの語は一般的な意味としても用いられるが、論文では専門的な意味で使われているので、専門用語としてリストに加えた。

6.1 専門用語の語彙リスト

頻度数55 structure
 頻度数52 DNA, 頻度数48 chain 頻度数44 base
 頻度数34 A (DNAのタイプを表すのでここでは専門語ととらえた)
 頻度数24 cytosine, model 頻度数23 fibre
 頻度数19 bond 頻度数18 coil 頻度数16 hydrogen
 頻度数13 sequence 頻度数11 phosphate replication
 頻度数10 nucleotide polynucleotide helical helice
 頻度数9 adenine guanine mythl crystallographic
 頻度数8 protein, thymine 頻度数7 B PH helix
 頻度数6 tyimine phage 頻度数5 adenine purine
 頻度数4では molecular pyrimidine plectonemic paranemic virus
 頻度数3 chromosome glucosidic meridian stereochemical hydroxyl
 注 hydroxyl 水酸基
 頻度数2 adenine bacteriophage electron enzyme histone lattice
 monomer mutation particle polymerigation potamine T titration
 titratable viscosity

注: lattice 「理」結晶格子

mutation 突然変異 pharge = bacteriophage

viscosity 「理」液体の粘度、粘性

頻度数1では、angstrom alkali birefringent birefringence bacterium
 carrier cytologist compound diffusion dichronism di-ester dyad dyplomerization
 diffraction elment glucoside infrared internucleotide ionizable ionization meridional
 microscope microscopy MW
 nitrogenous nucleotic nucleic nucleoprotamine physical-chemical physico-chemical
 para-crystalline phosphoryl paramenic plectonemic plectnemicly polypeptide py-
 rimidine sedimentation sodium
 stereochemically thymus tautometoric topological ultraviolet

註: birefringent 複屈折 carrier 保菌者 単体 cytologist 細胞学者

compound 化合物 diffusion 光り、ガスなどの拡散作用

dichroism 2色性 dyad ダイアド 「化」2価元素

diffraction 光り 音波 電波の回折 MW molecular weight

hydroxyl 「化」水酸基 microscopy 顕微鏡検査

meridional 子午線の sedimentation 沈降 sodium ナトリウム
 topological 「数」位相幾何学 地勢学

6.2 頻度数順に専門用語は何語あったか

6.1で述べた専門用語の異なり語の頻度数にしたがって、述べ語数を数えてみると、上記のように専門用語には合計100の異なり語があり、その述べ語数は636語である。

専門用語の異なり語とその述べ語数は下記の通り。

頻度数	55	52	48	44	34	24	23	19	18
異なり語	1	1	1	1	1	2	1	1	1
述べ語数	55	52	48	44	34	48	23	19	18
頻度数	16	13	11	10	9	8	7	6	
異なり語	1	1	2	4	4	2	3	2	
述べ語数	16	13	22	40	36	16	21	12	
頻度数	5	4	3	2	1				
異なり語	2	5	5	15	44	合計	100		
述べ語数	10	20	15	30	44	合計	636		

6.3 内容語と全体語数に対する専門用語の比率

内容語が2435語であったので、専門用語総数636語は、その26.1%に該当する。内容語ばかりでなく機能語も含めて、全体語数は5130であり、上述の専門用語は全体語数に対して12.4%にすぎなかった。専門用語はかなり限られている。

7. JACET8000の語彙リストを基準にすると

7.1 JACET8000の8レベルとは

JACET8000の語彙分類によれば、Level 1は中学校、Level 2は高校初級、Level 3が高校教科書および英検2級と定義されている。さらに解説によれば

Level 4 大学受験、大学一般教養の初級レベルに相当します。英語の単語力の有無を問われ分岐点です。

Level 5 大学受験、大学一般教養レベルに相当する。英検準一級レベル、TOEICでは400~500点がめざせます。少しずつ各専門領域の単語が増えてきます。

Level 6 英語を専門としない大学生やビジネスマンが目指すべきレベルである。英検では準1級、TOEICでは600点レベルに相当します。

Level 7 英語専門の大学生や英語を仕事で使うビジネスマンが到達目標とするレベルです。英検1級やTOEICでは95%以上の単語をカバーしています。

Level 8 日本人学習者の一般的な単語学習の最終目標です。あとは関連領域の専門用語

を増やすだけです。

このように8段階に分類されている。

7.2 JACET8000のLevel 5以上の単語の分類

以上のように8段階に分類され、Level 5が大学受験の語彙、TOEIC400～500点を指す語彙である。大学生が対象であるのでこのセクションではLevel 5を基準にして、Level 5以上の単語がどれくらいあるのか、すなわち大学受験以上の語彙がどれくらいなのかを分類してみた。

各頻度数別に、分類番号5000以上の単語をリストアップした。頻度数順に単語を分類したが、それぞれ5000以上、6000以上、7000以上の語に分類し、該当する場合はJACET8000の分類番号を記載した。派生語が分類番号を持っている場合には、派生語の番号に従って分類した。分類表にない単語があるので明記した。

頻度数18では 5825 coil, 頻度数10では7574 backbone, 頻度数9では5533 complementary, 頻度数8では6876 specificity, 頻度数5では 7572 postulate, 頻度数5で分類のない語はreflexion, 頻度数3では5495 complimentary, 6371 tangling, 7244 template, 7942 genetical, 頻度数3で分類のない語はperpendicular

頻度数2の語について

5000～の語は、5410 alternate, 5642 restrictive, 5747 plausible, 5712 inference,
5718 analytical

6000～の語は、6163 spice

7000～の語は 7049 linkage, 7439 regularity, 7978 duplicate, 7978 duplication

分類のない語は、humidity, mould, precursor, replication, titration, titratable

頻度数1の語について

5000～の語は、5092 simplicity, 5156 manipulation, 5348 rotation, 5349 residue,
5366 tilt, 5375 complement, 5436 puzzling, 5459 irrespective,
5512 ascertain, 5528 ambiguous, 5712 inference, 5714 symmetrically,
5813 superficially, 5884 displace

6000～の語は、6123 variance, 6234 isolate, 6254 microscopic, 6614 tentative,
6381 deduce (deduction), 6402 physiological, 6525 qualitative,
6600 geometrical, 6612 extraction, 6614 tentative, 6674 intriguing,
6793 satisfactorily

7000～の語は、7204 sperm, 7527 stable, 7791 estimation, 7943 inaccessible,
7978 unduplicated (duplicationとして分類)

分類のない語は、anomalous, apposition, asymmetrical circumference, crystalline, dis-

proof, equatorial, fabulous, imperative, insuperable, interlace, intermediary, intertwine, perpetuate, purify, recourse, reversible, superfluous, surmount, steric, symposium, synthesize, tabulate, thread, unbranch, upgrade, untreat, zipper-like

以上によればLevel 5以上の異なり語は86語あり、そのうち分類5000～5999の語は22語、分類6000～6999の語15語、分類7000～8000の語13語、分類なしが36語である。延べ語数は次の表の通り161語であった。*はJACET分類にない語を示す。

頻度数	18	10	9	8	5	5*	3	3*	2	2*	1	1*
異なり語	1	1	1	1	1	1	4	1	10	6	31	28
延べ語数	18	10	9	8	5	5	12	3	20	12	31	28

8. 語彙分析結果の考察

8.1 JACET8000リスト分類ごとのカバー率

総語数5130語のうち内容語は2435語なので、機能語は2695語になる。機能語リストの統計は本稿では算出していないので、内容語2435語について考えると、JACET8000リストによるレベル4までの語が1640語(67.4%)、レベル5以上の一般語が161語(6.6%)、専門語が634語(26%)である。大学受験、大学初級までの単語で内容語の67.4%をカバーしていることがわかった。機能語は算出していないのだがほとんど高等学校レベルで学習してある。機能語を含めて考えると、大学1年生にとって学習していない単語は、レベル5以上の語と専門語だけである。未修語は634語と161語をあわせて795語だけである。全体5130語のうち15.5%だけ知らない単語があることになる。知っている語が85.5%という結果になった。

8.2 学生のアンケートとの対比

学生アンケートによると、大学受験の語彙で科学論文を読みこなせると思うかどうかについては、思うと答えた学生は3名9%、29名91%は思わないと答えている。

受験の英語で論文語彙を何パーセントぐらいカバーしていると思いますかについては20～30%と答えたのは9名28%、30～40%が7名22%、40～50%は同じく7名22%、50～60%は1名3%、60～70%は5名16%、70～80%はゼロ、80～90%は3名9%であった。72%の学生が50%以下と回答している。

分析結果によれば受験語彙で85.5%カバーしているとでているが、アンケート結果では70%以上と答えているのは3名9%であった。ほとんどの学生が受験語彙ではどういサイエンスの論文は読めないと思っているのである。この意識のギャップを埋めるためにこういった語彙分析は有効であろう。数値があまりに明快にできたのが不思議ともいえる。一方、萩原氏はJACET8000の語彙だけでは、どうい生命科学の文献は読みこなせない

と述べているが、少なくともこの1篇の論文に関してはかなり高い語彙のカバー率であったと思う。読解だけでなく、リスニングにもスピーキングにも、ライティングにも語彙コーパスは活用できる。

9. まとめ

頻度数が最も高かったのは、structure, 次がDNA, 第3位はchain, 4位はA (DNAタイプを表す) 5位cytosine, 6位fibre, 7位 bond, suggest, 8位はcoil, model 第9位hydrogen, form, 10位only, pair 11位はoccur, 12位はfigure, group, single, X-ray, sequence 13位はdifficulty, specific, sugar, together 第14位はbackbone, helical, helice, nucleotide, polynucleotide と続く。全体語数5130語のうち内容語は2435語。そのうち専門用語は634語、JACET8000Level 4 以下の語は1643語、Level 5 以上が161語であった。JACET 8000level4までの受験語彙と大学初級語彙で85.5%をカバーしているという結果がでていますが、学生は32名中3名は、受験語彙で80~90%は論文の語彙をカバーしていることを知っているが、23名が50%以下と答えるほど、学生はとうてい読めないと思っている。この意識のギャップを埋めるために日頃の指導の一助として役立てたいものだ。

References

- 1 “The Structure of DNA” Watson—JD et al.
Cold Spring Harbor symposia on quantitative biology 1953 : 18 : 123-131
- 2 大学英語教育学会基本語リスト JACET8000 List of Basic Words
大学英語教育学会 基本語改定委員会編 2003年
- 3 JACET 8000 英単語 相沢一美、石川慎一郎、村田年編 桐原書店2005年
- 4 Taylor, C.V. 1979 “English of High School Textbooks”.
Canberra: Education research and development committee (ERDC) report No 18
- 5 Science World vol 2 朝日出版 2005年
- 6 Science Spectrum 先端科学技術ピックアップ 金星堂 2006年
- 7 生命科学分野で使用される英語語彙の分析：頻出語彙リストを使用した複数のコーパスの比較 萩原明子 第46回JACET全国大会 口頭発表 2007年